



CODATA-RDA Research Data Science Summer Schools

Vision, Information and Progress, Summer 2016

Objectives and Need

Contemporary research – particularly when addressing the most significant, transdisciplinary research challenges – cannot effectively be done without a range of skills relating to data. This includes the principles and practice of Open Science and research data management and curation, the use of a range of data platforms and infrastructures, large scale analysis, statistics, visualisation and modelling techniques, software development and data annotation. We define ‘Research Data Science’ as the ensemble of these skills.

Research Data Science skills are common to all disciplines and training in ‘Research Data Science’ needs to take this into account. For example, all disciplines need to ensure that research is reproducible and that provenance is documented reliably and this requires a transformation in practice and the promotion of the necessary culture, practice and skills.

A strategic priority shared by [CODATA](#) and the [Research Data Alliance](#) is to build capacity and to develop skills, training young researchers in the principles of Research Data Science. It is also important that Open Data and Open Science benefit research in Low and Middle Income Countries and do not result in even greater inequalities in research and scientific output. On the contrary, it has been argued that the ‘Data Revolution’ provides a notable opportunity for reducing research inequality in a number of respects. For this reason, particular attention is paid to the needs of young researchers in LMICs.

Mission of the CODATA-RDA Research Data Science Summer Schools

The CODATA-RDA Research Data Science Summer Schools will:

- address a recognised need for Research Data Science skills across disciplines;
- use and adapt existing materials to create an accredited curriculum that is more than the sum of its parts;
- provide a pathway from a broad introductory course for all researchers through more advanced and specialised courses;
- be reproducible: all materials will be online with Open licences;
- be scalable: emphasis will be placed on Training New Teachers (TNT) and building sustainable partnerships;
- pay particular attention to the needs of young researchers in LMICs.

Introductory School in Research Data Science

The introductory school will provide a bedrock of introductory material, common to all research disciplines, and upon which more advanced schools can build. The schools are primarily targeted at post graduate students, but early career researchers are also admissible. The schools provide skills that are of use to researchers and to other role involved in the research enterprise.

The introductory school is designed to run for two weeks and comprises six key components:

1. Principles and Practice of Open Science
2. Research Data Management and Data Curation
3. Software Carpentry
4. Data Carpentry
5. Infrastructures
6. Visualisation
7. Analysis

Progress and Resourcing for the First Introductory School

The first full introductory school will take place from 1-12 August 2016 at the [Abdus Salam International Centre for Theoretical Physics \(ICTP\)](#) in Trieste, Italy.

The venue, the Adriatico Guest House, is a delightful self-contained site, overlooking the sea, where there is accommodation, lecture halls, a terminal room and a canteen. The accommodation and subsistence is heavily subsidised. Optimal numbers for the smooth working of the school at ICTP will be 60-90 students. We received 326 applications for the first Introductory School of Research Data Science!

Reproducing the School: Fellows as helpers and training new tutors

We are putting in place mechanisms to ensure that the schools are as pedagogically effective as possible and that they can be easily reproduced elsewhere. To achieve this, we will run a Fellows programme, providing a cohort of students with special responsibilities. Fellows will help with the delivery of the schools, assisting their peers, and develop sufficient expertise in one of the six components to act as tutors for future schools.

An International Network and Advanced Schools

A number of organizations throughout the world have expressed an interest in running the introductory school in 2017 and 2018. Our objective is that as many as possible future hosts can be trained at the Trieste school in August 2016.

Our vision of further schools dealing with advanced and more discipline-specific and specialized data science skills. For example, in partnership with [CERN](#) we are preparing an advanced 'big data' school and in collaboration Scinecwe are developing a specialized school in the life sciences. The first delivery of these schools is planned to follow a second introductory school at ICTP in summer 2017.

Accreditation towards a Data Science Qualification and Business Model

The longer term vision is of a process for accreditation which would allow these short courses to contribute towards credits in an equivalent qualification.

In the short term we are building funded support for the first school and the follow-on events in 2017. In the medium term, we are working with colleagues at the Cologne Business School on a business model and business plan to sustain the initiative.

Further Information

Details about the schools can be found at <http://indico.ictp.it/event/7658/> and <http://www.codata.org/working-groups/research-data-science-summer-schools>